DØ Algorithms and Report from D0reco Task Force

Qizhong Li
Fermilab

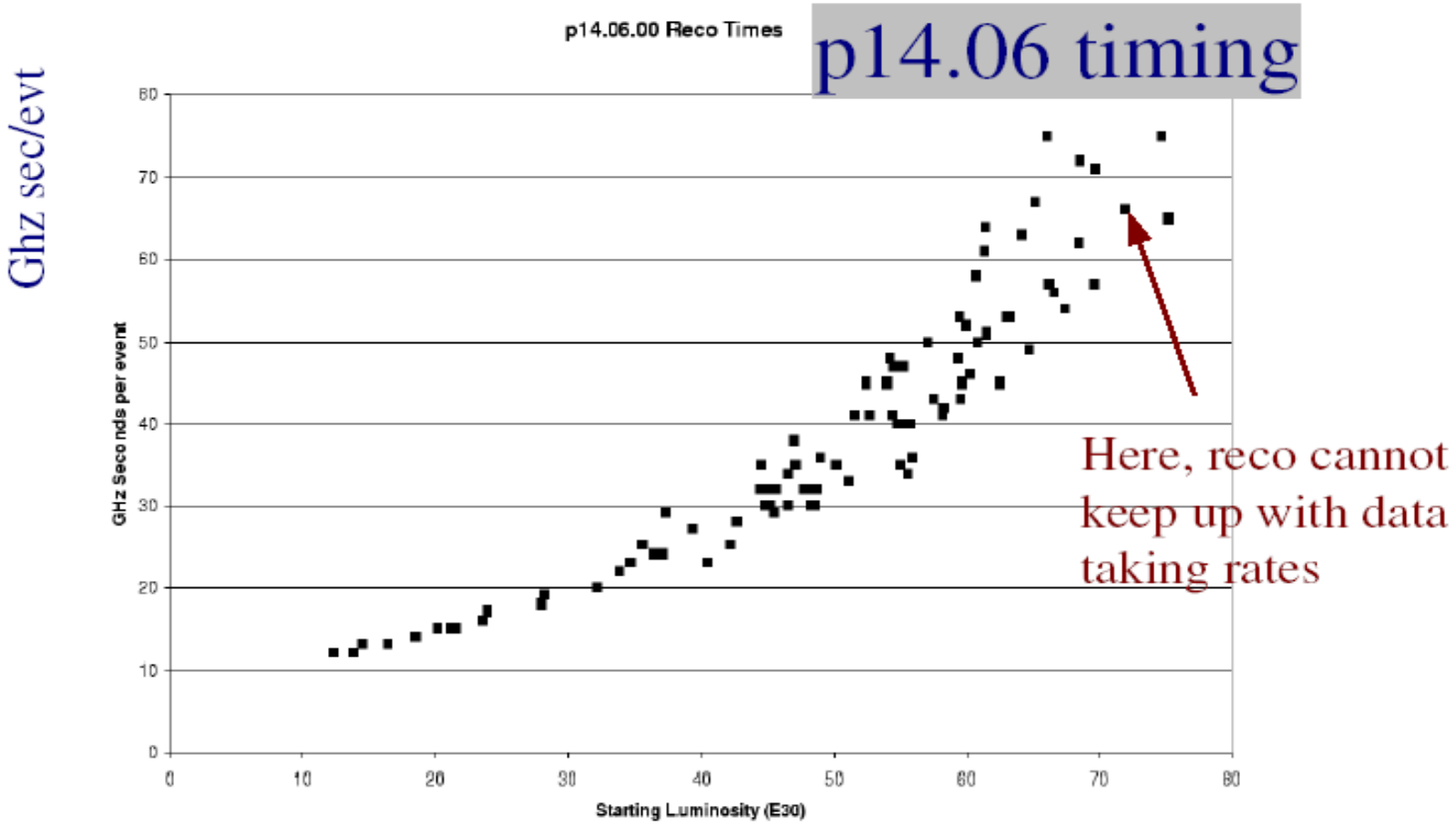
Run II Computing Review, September 2005

Outline

- Report from D0reco Task Force
  - D0reco speed up effort
- D0 algorithms
  - Focus on detailed understanding of detector systematics
  - Preparation of full 1fb^{-1} for quick analysis
  - Development of "second-generation" algorithms
  - Preparation for upgrades, higher luminosity
- Conclusions

P14 D0reco CPU problem

- Sept. 2004 with the higher luminosity, CPU problem became serious.



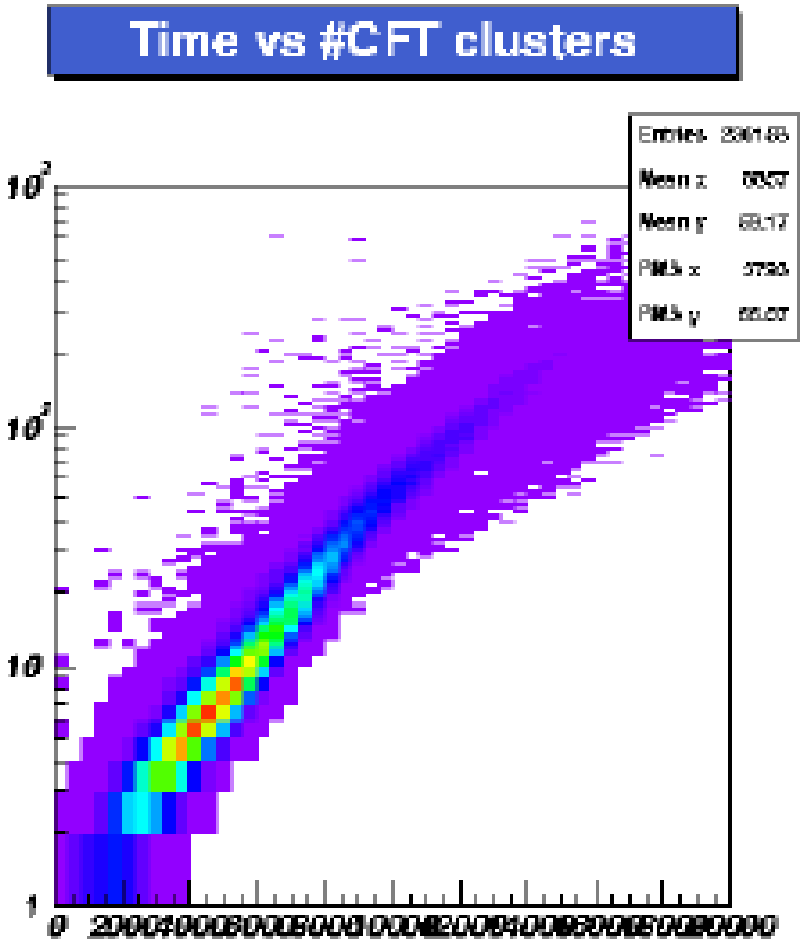
D0reco Task Force

- Charge: Speed up D0reco and reduce the slope to luminosity.
- Chair: Qizhong Li
- CD experts:
  - Philippe Canal, Jim Kowalkowski
- D0 tracking experts:
  - Guennadi Borissov, Mike Hildreth
- From physics groups:
  - Supriya Jain, Marine Michaut, Venkat Kaushik
- C++ code experts:
  - Scott Snyder, Paul Russo
- Executable managers:
  - Qizhong Li, Laurent Duflot, Suyong Choi
- Many people helped: calorimeter + cps experts, WZ and B physics group, …

Task Force's Efforts

- CD experts wrote tools to study CPU in details.
- Analyzed hot spots in D0reco CPU with higher luminosity data:
  - Tracking used 80% of D0reco CPU
    - AA tracking used 60%
    - HTF tracking used 20%
  - Calnada package used 3%
  - CPS unpacking used 3%
- Studied CPU time dependency with:
  - CFT occupancy
  - Number of vertices
  - Track reconstruction p_T threshold
  - Different triggers
  - …

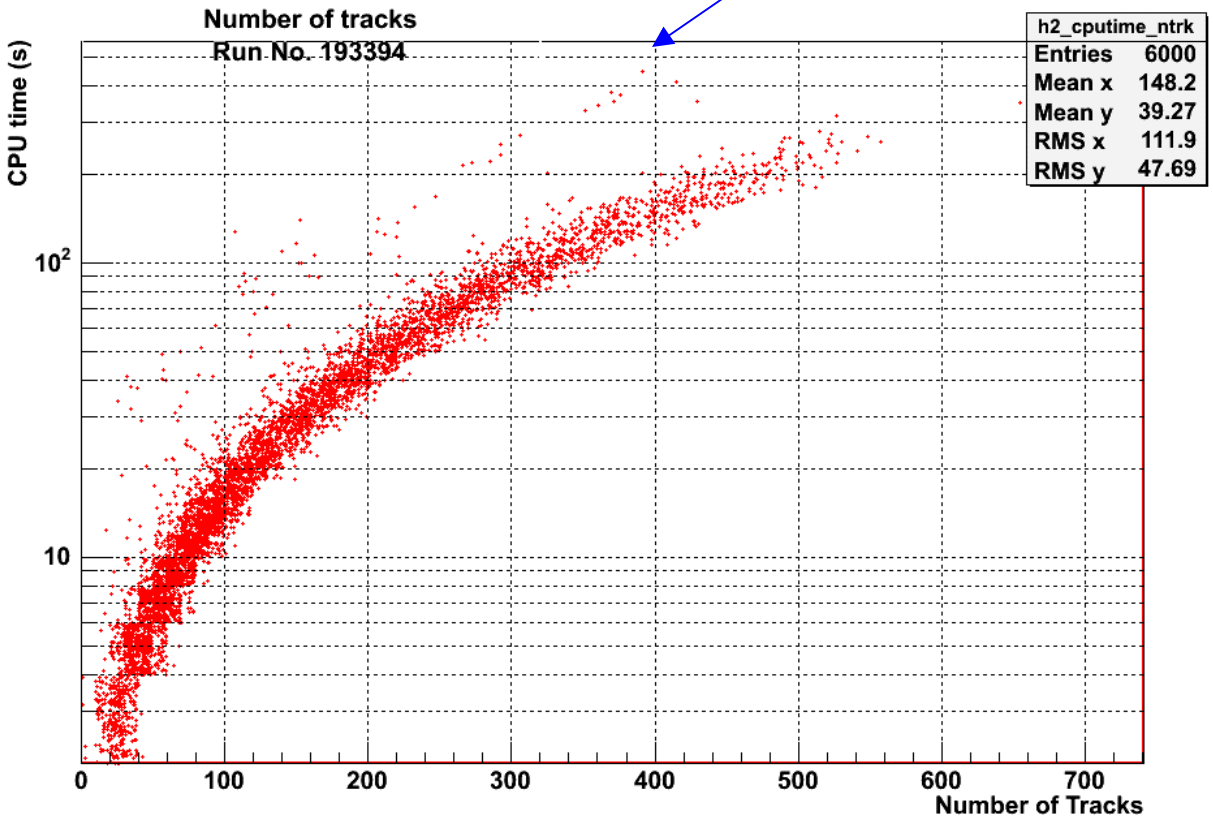
CFT Occupancy



- CPU time strongly correlated with CFT occupancy as expected

CPU vs Number of Tracks

A small fraction of events use a lot of CPU



CPU Speed Up Results

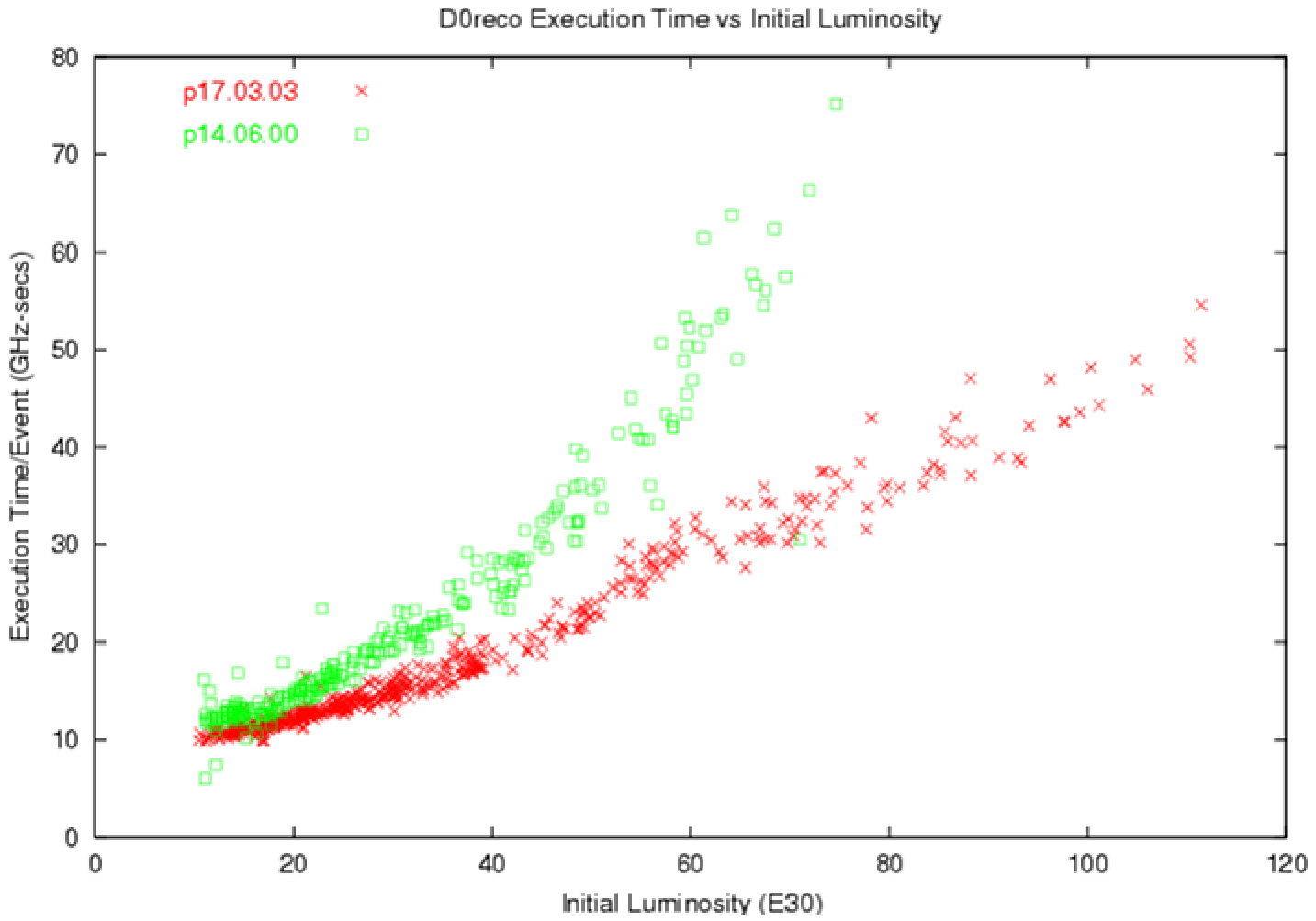
- ~15% CPU savings from improved code:
  - sin and cos calculations
  - matrix calculations
  - array access
  - lower level code improvements
- ~15% CPU saving from improved AA package:
  - Use big cluster instead of looking at each individual fiber.

- AA now no longer using most CPU
- Rewrote calnada and cps unpacking, no longer as hot spots.

The higher the luminosity, the more CPU savings!
With no efficiency loss, no effect on physics!

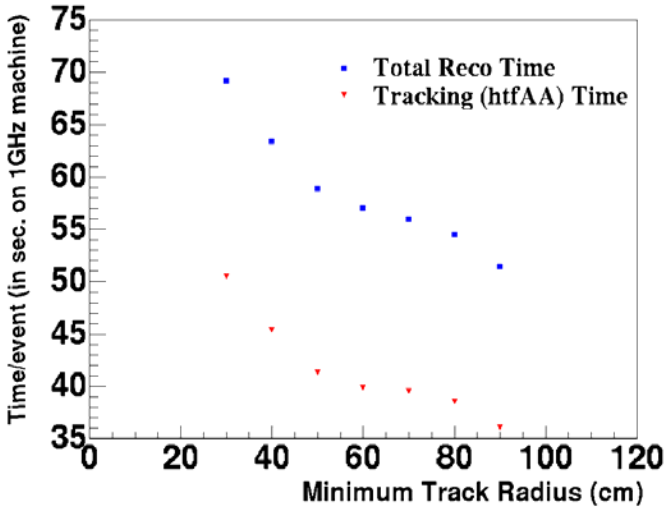
High Luminosity Improvement

- Improved D0reco CPU and less dependence on luminosity!



More improvements available for future

- The task force studied many possible speed-ups, not all of which were implemented.
  - Only those with no effect on efficiency were implemented.
- Developed several simple methods for more CPU savings to be used in future (with a few percent efficiency loss):
  - Not use the inner most layer as tracking seed;
  - Increasing track reconstruction p_T threshold.
- More complicated algorithmic changes are also possible.
- CD experts delivered a 20 page technical recommendation on how to improve AA tracking in future.

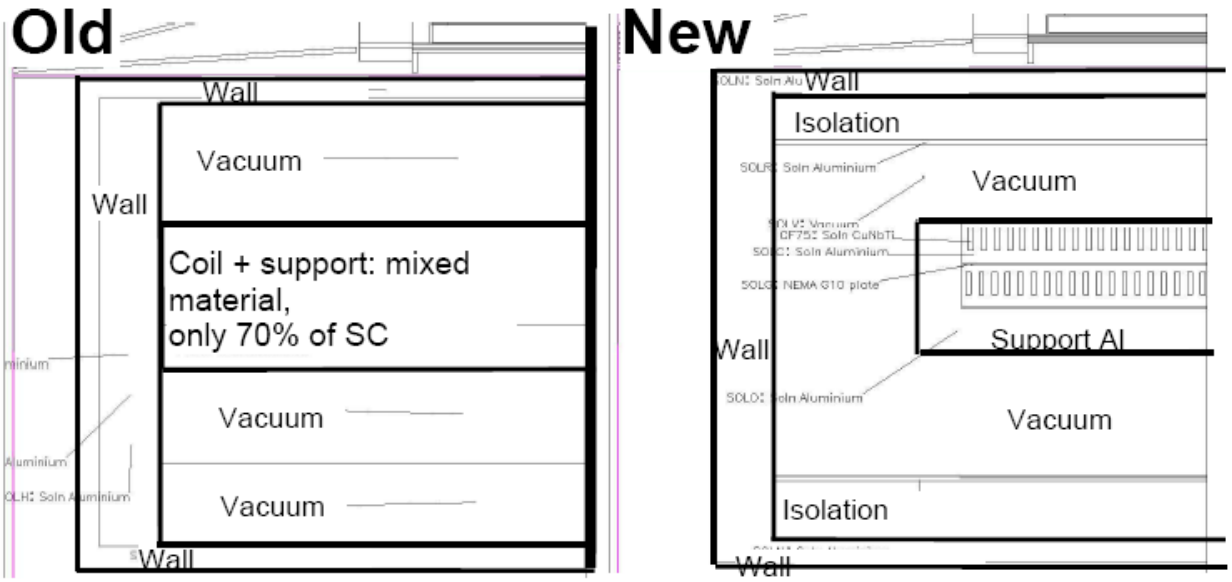


D0 Algorithms: Reaching Maturity

- Focus on detailed understanding of detector systematics
  - much improved description of detector material in MC and D0reco
  - improved simulation of tracker performance
  - zero-bias events used for noise & occupancy simulation
  - first full calorimeter calibration (EM and Hadron calorimeters)
- Preparation of full $1 fb^{-1}$ for quick analysis
  - reprocessing of entire dataset with above improvements
- Development of "second-generation" algorithms
  - Neural net b-tagging, new vertexing, etc.
- Preparation for Upgrades, higher luminosity
  - track timing optimization
  - algorithm modification for Layer 0 Silicon, CFT Timing, Triggers

D0 MC Improvements

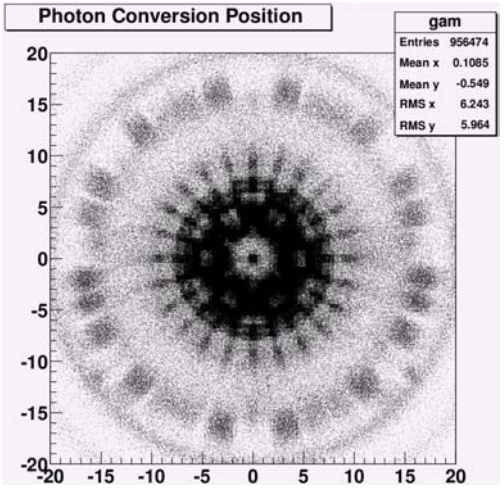
- Calorimeter Material Description:
  - Modification to Calorimeter/Cryostat/Solenoid

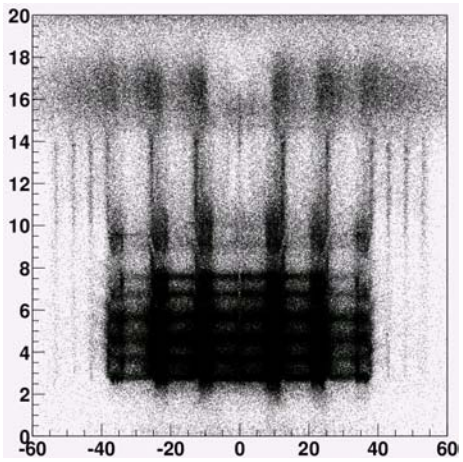


- added 10% X_0 more material, better description
  - MC Z mass resolution degrades by 20% - closer to data

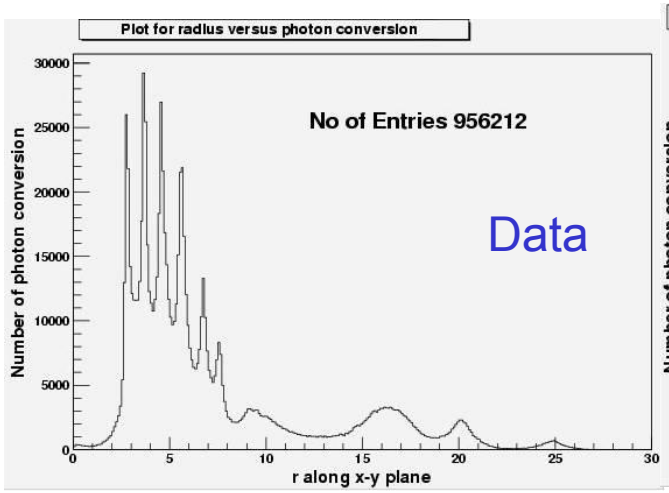
Tracker Material

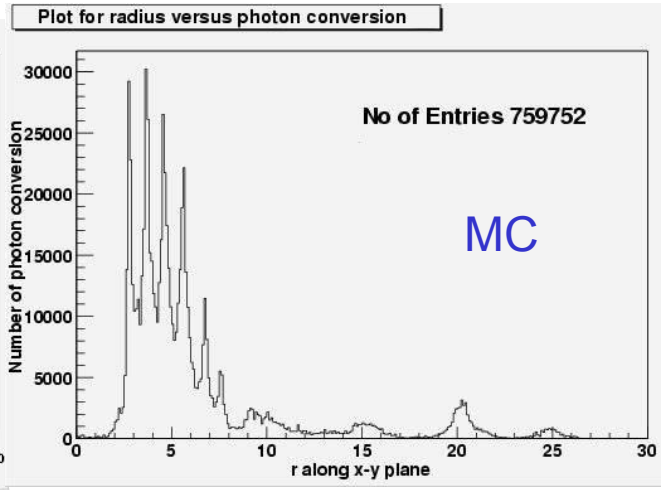
- Complete revamping of material in SMT volume
  - more detailed consideration of installed material
  - verification with photon conversions





- same material put in track fit

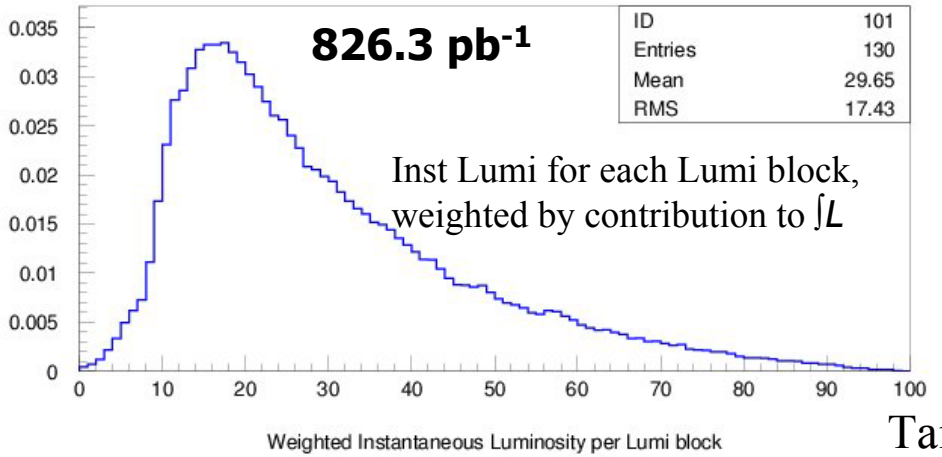




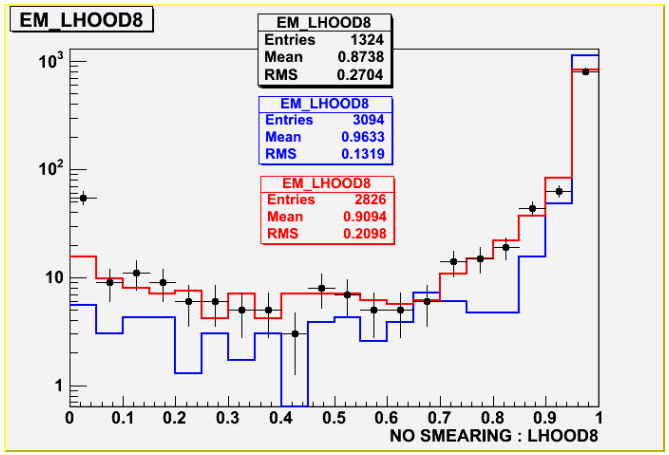
photon conversions vs. radius

Zero-bias in MC Simulation

- Overlay of zerobias events on top of MC hard scatter to simulate detector occupancy, noise...
  - one zerobias event per MC event
- Taken randomly from Run II lumi profile:

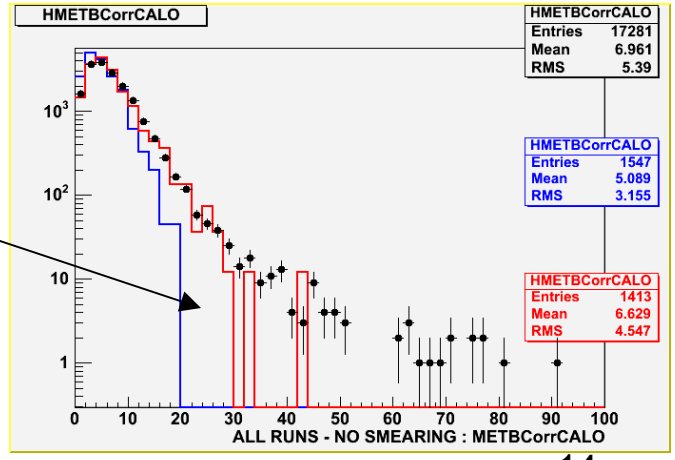


Z→ee events



EM likelihood

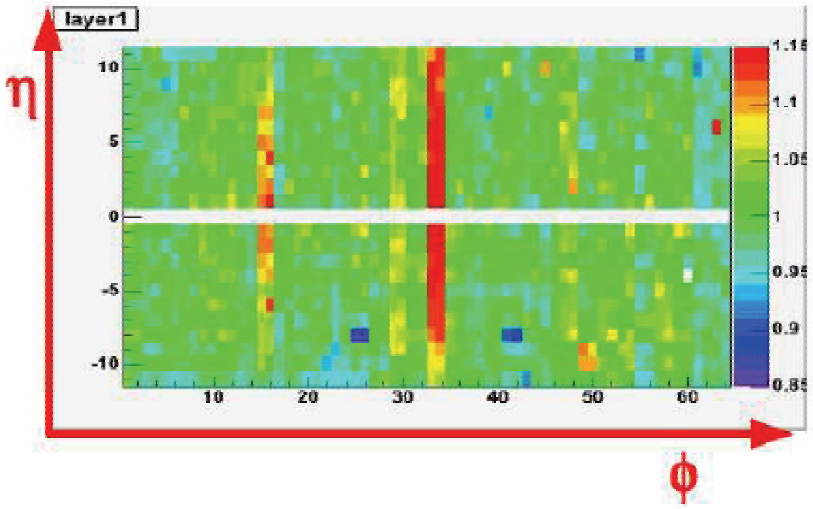
Tail never previously described by any MC
black: data
blue: Old MC
red: New MC

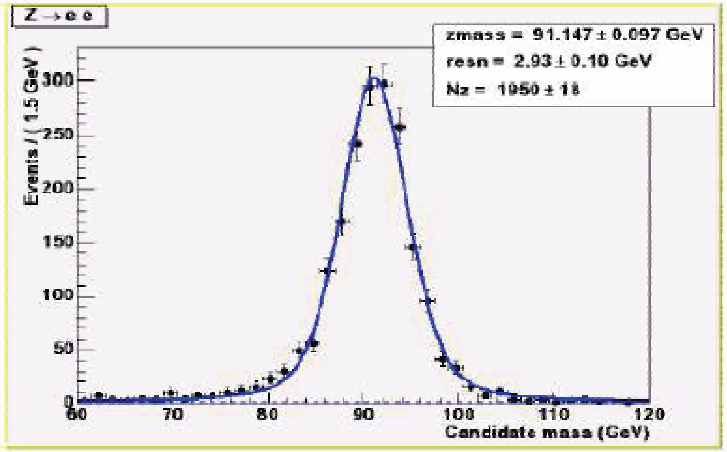


Corrected MET

Calorimeter Calibration

- determine multiplicative constant for each cell in the calorimeter to maximize energy resolution
- EM Calibration: measured completely from data
  - finished early 2005
  - response normalized in phi at fixed eta with special calibration sample
  - inter-eta calibration fixed using Z mass



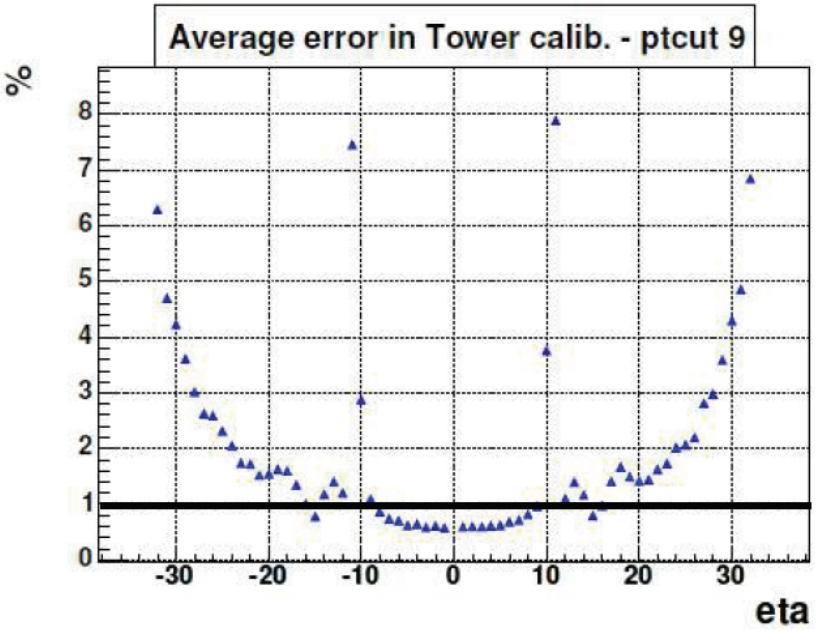


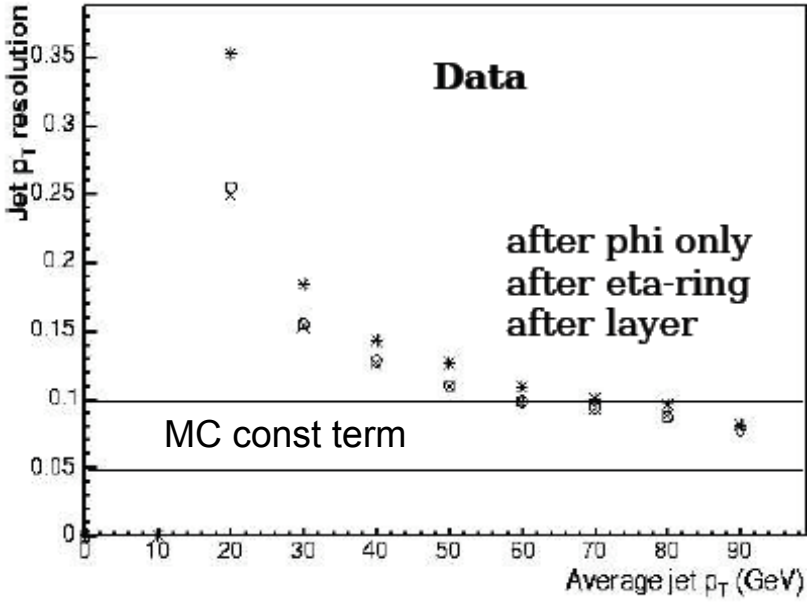
Detector resolution improved from 3.35 to 2.9 GeV

13%!

Calorimeter Calibration: HAD

- Finished in July
  - response normalized in phi at fixed eta with special calibration sample
  - inter-eta calibration fixed using di-jet data
  - Errors in Central Cal < 1%
    - ICD, high-η region limited by sample statistics
  - 5% rms variation of constants in each layer
  - comparisons with MC underway to understand asymptotic resolution





Preparation of 1fb^{-1} Dataset

- Full reprocessing of all Run II data with latest D0Reco release (p17)
  - take advantage of Calorimeter EM calibration
  - other improved algorithms and new algorithms
- "Fixing" pass for all data to apply HAD calibration, improved material description in track fitting.
  - uses information stored in TMB++ (thumbnail) data format
  - track re-fit from stored hit information
- Developed RecoCert program:
  - being used for D0reco and "Fixing" verification
  - being used for monitor data quality
    - runs on the production farm for every event

Preparation of 1fb^{-1} Dataset (Cont.)

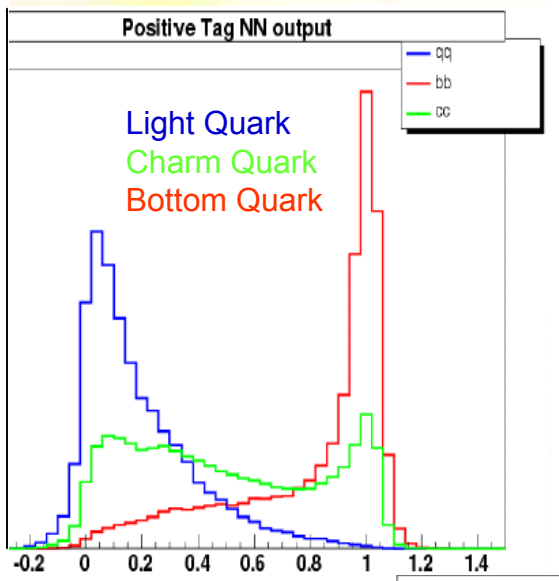
- Common Analysis Format (CAF):
  - Root-based micro-DST, common for all Physics groups
  - common object-selection, trigger-selection, normalization tools
  - simplify, accelerate analysis development
- Developed framework (CAFÉ) to analyze CAF formatted data
- Developing common analysis tools (CAF_UTIL) for particle ID selectors, corrections, systematic tool and efficiency tool.

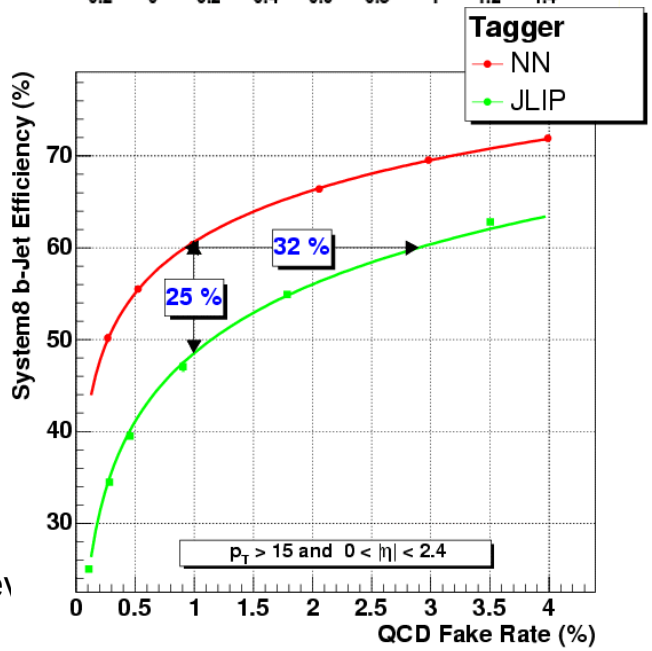
Reprocessing of all data will be done by October 2005.

All Run II data (~1fb^{-1}) will be in the same version of reconstruction, same analysis format.

New Algorithms: B-Tagging

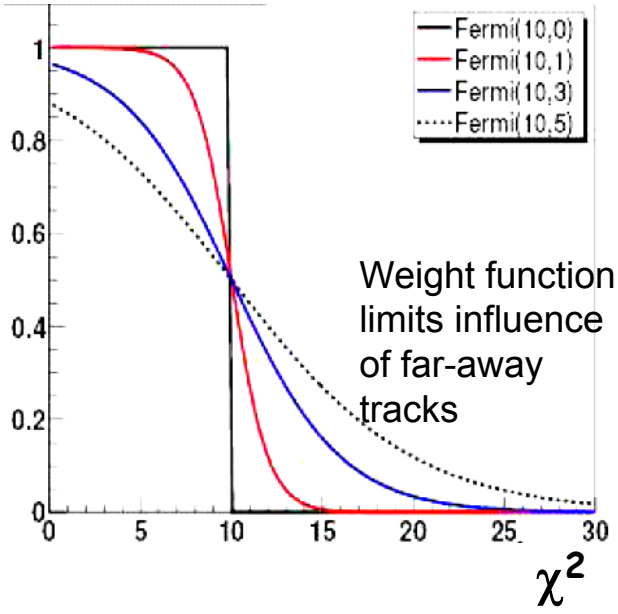
- Neural-net tagger, combining several simple taggers
- Achives a 25% increase in efficiency or a 32% decrease in backgrounds
  - still investigating biases, systematic errors
- Will be a boon to many statistics-limited analyses
- Hopefully a harbinger of more good things to come from combined tags, more advanced tagging algorithms





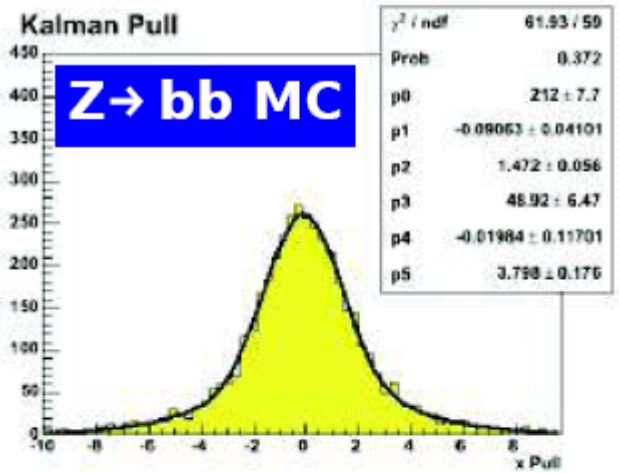
New Algorithms: Adaptive Vertexing

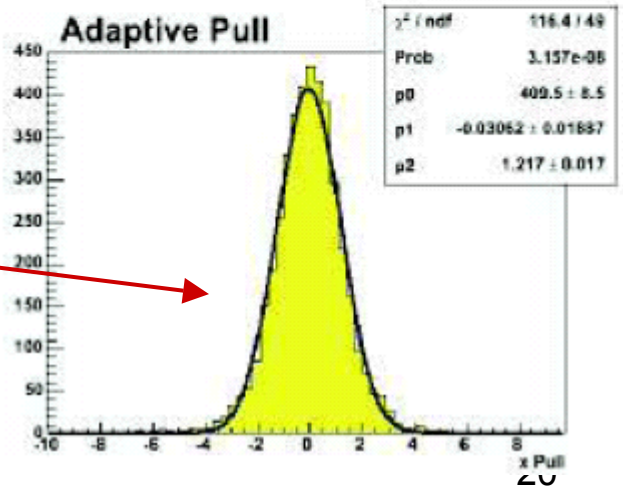
- Improved Primary Vertex resolution, less bias
  - benefits all analyses using lifetime information



Weight function limits influence of far-away tracks

Old Algorithm





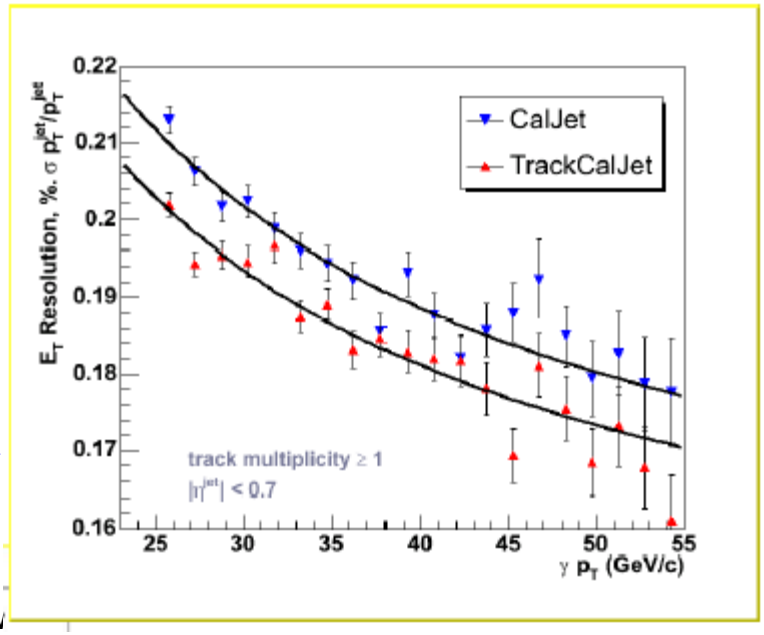
10-15% better resolution, fewer tails

- Also includes increased Z acceptance for primary vertex
  - SMT hits required only where necessary
- default vertexing in fixing pass

New Algorithms: Energy Flow

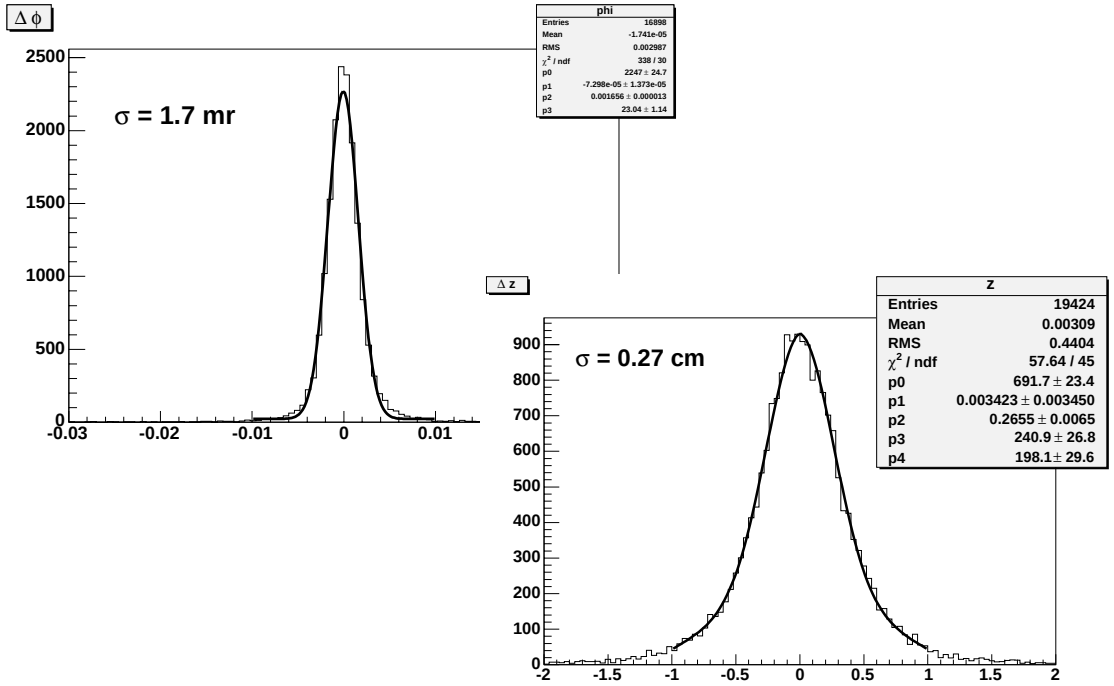
- TrackCalJet algorithm nearing maturity
  - Correction to jet energy based on average pion response and the individual track momenta
    - Individual track energies give more accurate correction to the energies deposited in the calorimeter
    - Corrections also made for tracks leaving the jet cone
- Gives significant improvement over calorimeter jets
  - 15% better on di-jet mass in $Z \rightarrow qq$
- Further improvements can be made by using a larger cone around the jet for including energy from tracks
- Will be default jet algorithm for jet energy scale
- Further improvements are coming from improved single pion studies

γ+jets MC



New in p17: Pre-shower and Photon ID

- Reconstructed CPS information present for the first time in p17.

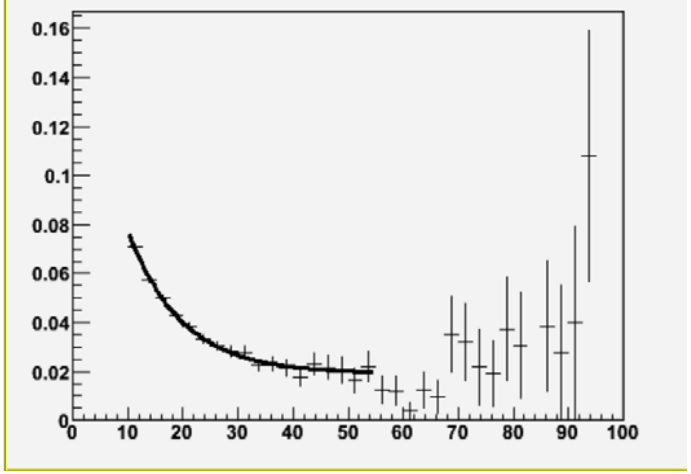


Δϕ and Δz between a CPS cluster and a track matched with an EM object.

- Photon ID is new.

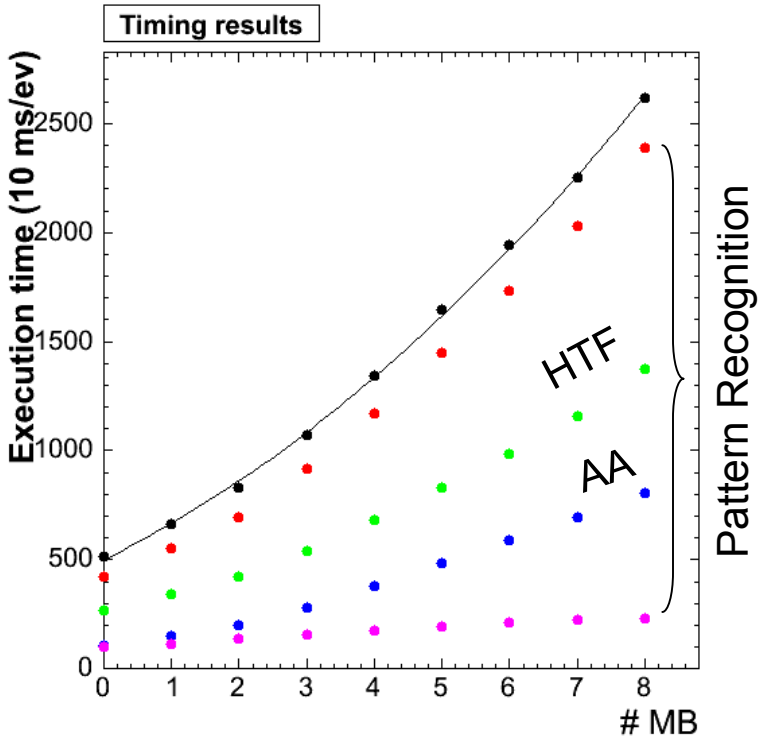
Photon ID incorporates tracking information by searching for hits along a road pointing from the event vertex to the EM cluster (Hits on the road method).

Misidentification rate vs p_T



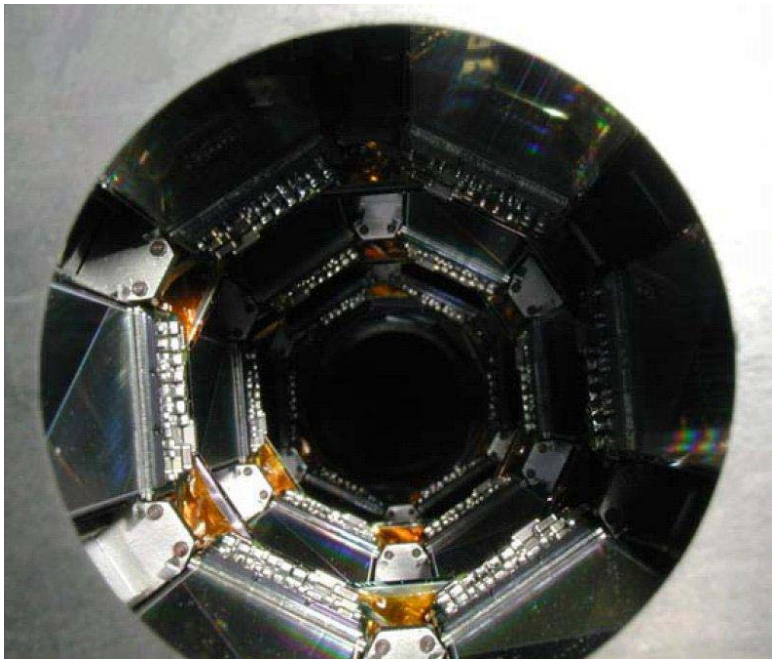
Future Timing Improvements

- Combined tracking algorithm
  - two components:
    - AA – road-based algorithm
    - HTF – histogram-based
  - AA was main focus of D0reco Task force
  - HTF now dominates CPU time
- Recent studies suggest another 20-30% time reduction in tracking timing may be possible by modifying HTF to concentrate on forward tracking
  - little loss in central efficiency
- Further structural mods to AA will be studied; should give further increase in speed

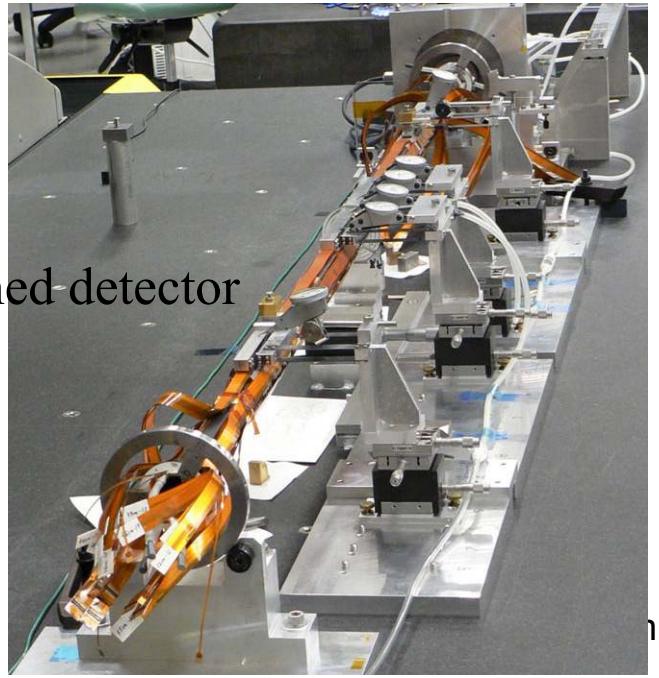


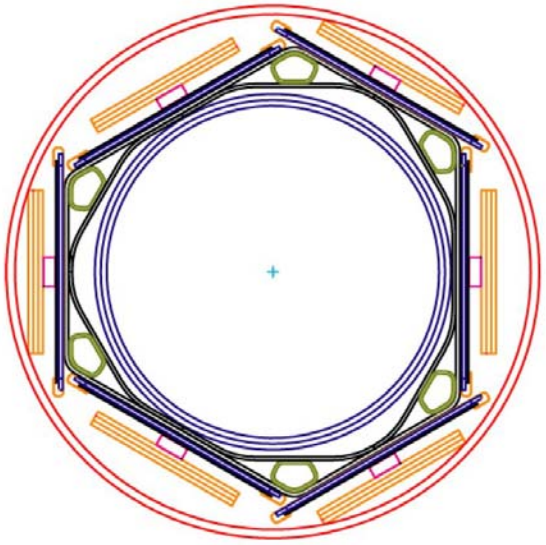
Preparing for Upgrades: SMT Layer 0

- fits inside current SMT on smaller beampipe
- software is ready for layer 0
  - GEANT geometry, digitization code, reconstruction geometry ready
  - incorporation of new hits into pattern recognition almost complete
- final material corrections TBD



finished detector





Preparing for Upgrades: AFEII-t

- Replacement of CFT electronics to cure many occupancy- and rate-dependent pathologies
  - will include timing information on each fiber (Trip-t chip)
  - simulation results indicate timing resolution of ~1.8 ns/fiber
    - ~36cm in z
    - consistent with measured resolution in lab tests
    - will be useful for improving speed of pattern recognition or reducing fake or mis-measured tracks at highest luminosities
      - only use hits consistent in z with track road
      - 40% speed-up seen in modified tracking version
- Aiming for summer/fall 2006 installation
  - algorithm development project to optimize use of timing info
  - calibration/monitoring effort will need to be developed

Trigger Upgrades: Related Software

- DØ Trigger upgrades include the replacement of most elements of the Level 1 Trigger system, improvements to Level 2 and Level 3.
  - Level 1 Calorimeter trigger and new Level 1 Calorimeter-Track match system include many new capabilities
    - clustering, tau algorithms, track-jet verification all at Level 1
    - complicated topological triggers possible
- Original simulation software used for system design now mature, fully integrated into the official DØ Trigger Simulation
  - work has passed from Upgrade Project to Trigger Studies Group
  - allows Level 2 and Level 3 code development
  - already used to design a preliminary trigger list using the new trigger objects
    - new trigger able to withstand 2E32 within Level 1+Level 2 rate budget
  - many studies ongoing

Conclusions

- Primary focus on understanding ultimate detector performance
  - minimize systematic errors for precision measurements
    - requires optimal treatment of data
    - requires excellent MC-Data agreement

In Parallel:

- preparation of 1fb^{-1} dataset
  - reprocessing, fixing, new analysis data format, new MC
  - will be ready for Moriond 2006 conference
- new second-generation algorithms to optimize performance
- preparation for higher luminosity, detector upgrades
  - focus on maintaining efficiency, minimizing reconstruction time

Many projects underway/approaching completion.